日本オペレーションズ・リサーチ学会
2004年春季研究発表会

SVMによる企業の倒産予測

01991675　香川大学　＊　尹 禮分　YUN Yeboon
01401604　甲南大学　　中山 弘隆　NAKAYAMA Hirotaka

1　はじめに

1990年代 Vapnik らにより提案されたサポートベクターマシン（簡単に SVM という）は近年パターン分類の方法として注目されている．SVM では，非線形写像によって元のデータを高次元の特徴空間に移し，線形分離可能な超平面を求める．さらに，マージン（各データから分離平面までの最短距離のことを言う）を最大化することによって未知のデータの対する予測の精度が高くなることが証明されている．そこで，本研究では SVM の一種であるトータルマージンアルゴリズムを用いて企業の倒産予測し，特に倒産企業に対する判別率が従来の SVM より高いことを示す．

2　サポートベクターマシン

与えられた学習データ

$$(x_1, y_1), \cdots, (x_\ell, y_\ell),\ x_i \in \mathbb{R}^n,\ y_i \in \{-1, 1\}$$

に対し、分類関数

$$f(\Phi(x)) = w^T \Phi(x) + b$$

を考える．ここで，Φ はある非線形写像である．

現実問題では，特徴空間に写像された学習データでも線形分離できないことが多く，しかも完全分離が必ずしもよいとは言えないことがある．そこで，スラック（slack）変数 (Fig. 1) が用いられた SVM が提案された．

$$\begin{aligned}
&\min_{w,b,\xi_i} \ \frac{1}{2} w^T w + C \sum_{i=1}^{\ell} \xi_i && \text{(S)}\\
&\text{s.t.} \quad y_i \left(w^T z_i + b \right) \geqq 1 - \xi_i,\\
&\qquad \xi_i \geqq 0,\ i = 1, \ldots, \ell,\\
&\max_{\alpha} \ \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} y_i y_j \alpha_i \alpha_j \langle \Phi(x_i) \cdot \Phi(x_j) \rangle\\
&\text{s.t.} \quad \sum_{i=1}^{\ell} y_i \alpha_i = 0, && \text{(S_D)}\\
&\qquad 0 \leqq \alpha_i \leqq C,\ i = 1, \ldots, \ell.
\end{aligned}$$

ここで，C はスラック変数に対するパラメータである．

さらに，著者らは正しく判別されたデータが分類超平面からどれだけ離れているかを表しているサープラス

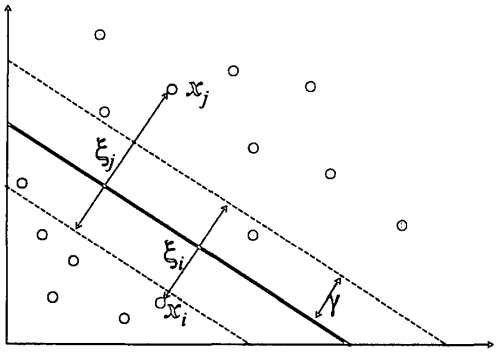

Fig. 1: slack Variables

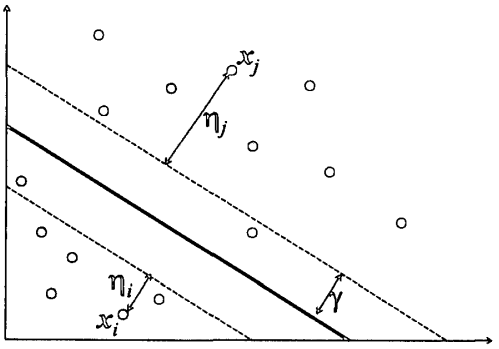

Fig. 2: surplus Variables

(surplus) 変数 (Fig. 2) を定義し，スラック変数とサープラス変数を同時に考慮するトータルマージンアルゴリズム (T)

$$\begin{aligned}
&\min_{w,b,\xi_i,\eta_i} \ \frac{1}{2} w^T w + C_1 \sum_{i=1}^{\ell} \xi_i - C_2 \sum_{i=1}^{\ell} \eta_i && \text{(T)}\\
&\text{s.t.} \quad y_i \left(w^T \Phi(x_i) + b \right) \geqq 1 - \xi_i + \eta_i,\\
&\qquad \xi_i \geqq 0,\ \eta_i \geqq 0,\ i = 1, \ldots, \ell
\end{aligned}$$

と、問題 (T) に対する双対問題 (T_D)

$$\begin{aligned}
&\max_{\alpha} \ \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} y_i y_j \alpha_i \alpha_j \langle \Phi(x_i) \cdot \Phi(x_j) \rangle\\
&\text{s.t.} \ \sum_{i=1}^{\ell} y_i \alpha_i = 0, && \text{(T_D)}\\
&\qquad C_2 \leqq \alpha_i \leqq C_1,\ i = 1, \ldots, \ell\ (C_1 > C_2)
\end{aligned}$$

を提案した．ここで，C_1 と C_2 はそれぞれスラック変数とサープラス変数に対するパラメータであり，$\langle \Phi(\boldsymbol{x}_i) \cdot \Phi(\boldsymbol{x}_j) \rangle \equiv K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ となる．本論文の実験では次のガウシアンカーネル関数を用いる．

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2}{2\sigma^2}\right).$$

3 倒産・非倒産の判別

実際のデータを用いて，2 節で紹介した従来の SVM とトータルマージンアルゴリズムにより倒産企業と非倒産企業の判別を行い，その結果を比較する．

東京商工リサーチから提供された 1999 年の財務データから，建設業の中小企業データを使用する．財務指標は，自己資本率，借入金依存度，売上高総利益率，売上高支払利息比率，総資本回転率の 5 つを選定する．全データの中，倒産企業数より非倒産企業数が極端に多いので，表 1 に示すデータ数で数値実験を行う．ランダムで抽出したデータに対する判別の正解率（10 回試行した結果の平均正解率）を表 2 と表 3 に示す．

表の結果からもわかるように，トータルマージンアルゴリズムによる正解率が，従来の SVM による正解率より全体的に優れている．特にデータ数が少ない倒産企業に対して，従来の SVM によりトータルマージンアルゴリズムがより正確に判別している．従来の SVM では正解率は高くても 30% ぐらいであるが，トータルマージンアルゴリズムでは，サープラス変数に対するパラメータ C_2 が大きいとき，正解率が上がる．トータルマージンアルゴリズムでは，誤判別データに関する測度として用いられるスラック変数だけではなく，正判別されたデータに関する測度のサープラス変数を同時に考慮しているため，よりいい結果が得られている。

表 1: データ

データ	学習		テスト	
	倒産	非倒産	倒産	非倒産
集合 1	50	300	100	100
集合 2	100	400	100	100

表 2: 従来の SVM による結果

(1) 倒産企業に対する正解率

(a) データ集合 1

C	1	10	100
学習	17.4	45.2	73.2
テスト	1.8	15.5	23

(b) データ集合 2

C	1	10	100
学習	23.5	51.1	69.3
テスト	9.9	25.9	25.2

(2) 非倒産企業に対する正解率

(a) データ集合 1

C	1	10	100
学習	99.9	99.4	99.47
テスト	99.8	96.4	90.6

(b) データ集合 2

C	1	10	100
学習	99.3	98.15	98.63
テスト	97.5	91.9	90

表 3: トータルマージンアルゴリズムによる結果

(1) 倒産企業に対する正解率

(a) データ集合 1

	$C_1 = 1$		$C_1 = 10$		$C_1 = 100$		
C_2	0.01	0.1	0.1	1	0.1	1	10
学習	18.6	85	62	95.6	95	100	93.8
テスト	2.7	74.5	30.9	89	53.9	82	89.6

(b) データ集合 2

	$C_1 = 1$		$C_1 = 10$		$C_1 = 100$		
C_2	0.01	0.1	0.1	1	0.1	1	10
学習	30.1	81.6	63.3	96	95.2	99.7	95.4
テスト	10.6	71.5	39.3	92	60.5	80.9	94.4

(2) 非倒産企業に対する正解率

(a) データ集合 1

	$C_1 = 1$		$C_1 = 10$		$C_1 = 100$		
C_2	0.01	0.1	0.1	1	0.1	1	10
学習	99.87	69.9	98.47	49.63	77.37	49.67	49.33
テスト	99.4	68	89.3	46.4	69.4	47.1	48.5

(b) データ集合 2

	$C_1 = 1$		$C_1 = 10$		$C_1 = 100$		
C_2	0.01	0.1	0.1	1	0.1	1	10
学習	98.67	71.17	96.15	43.63	75.1	46.55	40.13
テスト	96.5	66.4	88.6	42.9	69.2	43.8	36.6

参考文献

[1] C. Cortes and V. Vapnik, Support vector networks, *Machine Learning*, vol. 20, pp. 273–297, 1995

[2] B. Schölkopf and A.J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond-*, MIT Press, 2002.

[3] M. Yoon, Y.B. Yun and H. Nakayama, Using Total Margin in Support Vector Machines, Proceedings of IJCNN2003